International Journal of Innovative Research in Engineering and Management (IJIREM)
ISSN (Online): 2350-0557, Volume 13, Issue 5, Oct-Nov 2026
DOI: https://doi.org/10.55524/ijirem.2026.13.5.6
Article ID IRP-1794, Pages 47-53
www.ijirem.org

Artificial Intelligence and Machine Learning, for Thyroid Disease and Thyroid Disorder Prediction and Detection: A Comprehensive Review of Techniques Clinical Features, Applications, Explainability, Challenges and Future Directions

Roshani Koli[1] , and Hemangini Patel[2]

[1] Researcher Scholar, Department of Computer Science, Vanita Vishram Women's University, Surat, Gujrat, India
[2] Department of Computer Science, Vanita Vishram Women's University, Surat, Gujrat, India

Correspondence should be addressed to Roshani Koli; koliroshni28@gmail.com

Received: 26 Sep 2026; **1st Revision:** 8 Oct 2026; **2nd Revision:** 15 Oct 2026; **Accepted:** 21 Oct 2026;
Published Online: 28 Oct 2026

Copyright © 2026 Roshani Koli Li. This is an open-access article distributed under the Creative Commons Attribution 4.0 International License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

ABSTRACT- Thyroid problems like hypothyroidism and hyperthyroidism are some of the common hormone-related issues. Getting the diagnosis quickly is very important for proper treatment. Traditional ways of diagnosing these problems rely on checking how a person feels and doing blood tests, for TSH, T3 and T4. As more digital health data becomes available, artificial intelligence, machine learning, deep learning and data mining are being used to help predict thyroid diseases faster and better. This review puts together the developments in AI-based thyroid prediction. It covers the kinds of data sets that are used how the data is processed and which features are chosen, machine learning algorithms, ways to combine models and some methods that are inspired by biology. It also looks at deep learning models ways to create data when there is an imbalance and tools that make the AI easier to understand for doctors. Research from studies shows that models such as Random Forest, XGBoost Gradient Boosting, SVM, Decision Trees, KNN and different types of combined models can have very good prediction results. Some of these models have reported accuracy rates of over 98 to 99 percent when they use feature selection methods, like Featurewiz and SelectKBest and techniques to balance the classes like making more data or using GANs to create new data. More recent work has started to use ideas from nature to create optimization algorithms. It also uses types of structures like Kolmogorov–Arnold Networks. It uses methods that make results easier to understand such as SHAP and LIME. These changes help improve both accuracy and transparency. However there are still problems. These problems include data sets that're small and not very different from each other. There is also a problem with some categories having more examples than others. There is not testing done in different places or with different groups. The data sets are not the same in places. Some data is missing. There is not information about patients over time.. There is very little use of these methods, in real hospitals or clinics. Future research should focus on building clear, multimodal and long-term aware AI systems. These systems need to be tested outside the setting and must help doctors make steady real-world decisions, in thyroid care every day.

KEYWORDS- Artificial Intelligence, Machine Learning, Deep Learning, Thyroid Disease, Thyroid Disorder, Feature Selection, Random Forest, XGBoost, Ensemble Learning, Explainable AI(XAI), SHAP, LIME, Data Mining.

I. INTRODUCTION

Thyroid disorder detection using machine learning has become an area of research. Many studies compare algorithms and ensemble methods to improve how well models can predict thyroid conditions. One approach, called stacking ensemble achieved an F1-score of 0.9944 using a dataset with 9,172 records [1]. This shows how powerful machine learning can be when applied carefully. The link between thyroid-stimulating hormone (TSH) and thyroxine (T4) is key to understanding thyroid problems. A study with 110 patients showed that TSH and T4 levels move in directions. In patients with hypothyroidism the correlation was −0.82. In hyperthyroidism it was −0.78 [2]. These strong inverse relationships help clinicians spot thyroid function. Ensemble methods when paired with feature selection and strategies to balance classes have produced results. One study reached 99.72% accuracy and 100% sensitivity using a predictive model [3]. That means the model correctly identified all cases without missing any real positives. This is a step toward reliable diagnosis. Bio-inspired optimization methods have also been tested. These mimic natural processes like evolution or swarm behavior to fine-tune machine learning models. They help boost the performance of thyroid disease classification systems [4]. Deep learning has entered the field too. Advanced models like adversarial networks (GANs) and Kolmogorov–Arnold networks (KANs) have been used for classifying three types of thyroid diseases [5]. These models can learn patterns from data, which helps in spotting subtle differences. Explainable artificial intelligence (XAI) is another area. It makes machine learning models easier to understand, which is essential in healthcare. A Random Forest model achieved 91.42% accuracy. Included explainability features that help doctors trust the predictions [6]. This transparency is vital when models are used in clinical settings.

Some studies use Featurewiz-based feature selection to pick the relevant variables. This helps reduce noise and improve classification accuracy [7], [8]. By focusing on features models become faster and more accurate. Empirical studies have helped build thyroid disease classification systems.These evaluations guide the development of tools for diagnosis. Feature selection is also used to remove repetitive inputs. By focusing on the informative attributes models become more efficient and less prone to overfitting [9]. This improves both performance and reliability. Comparative experiments test machine learning algorithms—like logistic regression support vector machines and decision trees—to see how well they predict thyroid disease risk [10]. These comparisons help identify the effective methods. Machine learning is not used for diagnosis. Its also being explored for predicting how patients will respond to treatment. This expands its use beyond identifying illness to guiding care decisions [11]. Performance comparisons help researchers understand what works and what doesn't. They show the strengths and weaknesses of models in real-world scenarios [12]. This knowledge is essential for making choices. Many algorithms have been applied to thyroid disease classification. The goal is to automate the detection process so that doctors can get more consistent results [13]. This could help reduce errors and improve outcomes. Explainable AI is being studied in healthcare to make predictions more transparent. Clinicians need to know why a model made a decision especially when it affects health [14]. These builds trust in AI systems. Data mining techniques are used to find patterns in data. These methods support the development of tools that can analyse large datasets and identify thyroid disorders [15].

The UCI Machine Learning Repository offers datasets that researchers can use to develop, test and compare models. This shared resource helps standardize research and improve reproducibility [16]. Interactive machine learning systems have been proposed. These allow users to input data and get feedback helping with the early detection of thyroid disorders [17]. Such systems can be useful in clinics and remote areas.

Hybrid systems that combine methods—like adaptive neuro-fuzzy inference systems (ANFIS) K-nearest neighbors (KNN) and information gain—have been tested for forecasting thyroid diseases [18]. Combining strengths from approaches often yields better results. Comparative studies between KNN and Naïve Bayes have looked at classification methods. These studies help determine if older models still hold value when new ones are available [19]. Data mining is also used to find patterns in thyroid data. These approaches help create tools that can detect thyroid problems from patient records [20].

Overall the research shows a growing use of machine learning, ensemble methods, feature selection, optimization and explainable AI in thyroid disease prediction. Despite progress challenges remain. Missing data, categorical variables redundant features, class imbalance, model interpretability and generalizability are still issues [1], [3], [6], [9], [10], [12] Solving these problems is key to building models that're not only accurate but also reliable, interpretable and useful, in real healthcare environments.

II. LITERATURE REVIEW

The recent research shows a move toward big ensemble and hybrid methods. Zhong and Wang [1] examined machine learning and deep learning algorithms on 9,172 observations and reported that a stacking ensemble with XGBoost as the meta-learner produced the highest F1-score of 0.9944. I find XGBoost fascinating as XGBoost dominates the charts. Ali and colleagues [2] proved the foundation for computational prediction by showing strong inverse relationships, between TSH and T4 in hypothyroid patients (r = -0.82) and hyperthyroid patients (r = -0.78).. Teammates [3] tested 17 machine-learning models. Reported that XGBoost combined with SelectKBest feature selection and random oversampling achieved 99.72 % accuracy and 100 % sensitivity. Kesavulu and Kannadasan [4] introduced Particle Snake Swarm Optimization combined with Random Forest (PSSO-RF) reaching 98.7 % accuracy and surpassing a CNN-LSTM baseline. Group[5] merged adversarial networks with Kolmogorov-Arnold Networks and SHAP/LIME explainability attaining 98.68 % accuracy while identifying TSH as the most influential feature.

Earlier, in 2023 Hossain et al. [6] developed an AI framework for hypo- and hyperthyroidism reporting that Random Forest achieved 91.42% accuracy after feature-importance-based selection while SHAP and LIME confirmed TSH, T3, TT4 and FTI as key predictive features. Lim et al. [7] introduced Featurewiz-based feature selection for thyroid data showing that Random Forest accuracy improved from 98.79% to 99.45% (F1-score 0.99 AUC 0.97) after feature reduction.

Research in 2022 emphasized feature engineering and broader algorithm comparison. investigated machine learning combined with feature engineering for thyroid classification. Chaganti et al. [9] compared selective-feature approaches across machine-learning and deep-learning models, including Random Forest, Logistic Regression, SVM, AdaBoost, Gradient Boosting, CNN, LSTM and CNN-LSTM achieving accuracy of, up to approximately 0.99. Islam et al. [10] evaluated ANN, CatBoost and XGBoost on a Sick-euthyroid dataset with ANN achieving the accuracy of 95.87%.

Between 2016 and 2021 many studies set comparative standards. Aversano and colleagues looked at machine-learning methods to predict outcomes of thyroid disease treatment. Rehman and colleagues examined how many machine-learning algorithms perform when classifying thyroid disease. Salman and Sonuc compared machine-learning algorithms for tasks that classify thyroid disease. Pawar and colleagues discussed adding AI so that doctors could better understand machine-learning predictions, in health care. Begum and Parkavi used data-mining techniques to predict thyroid disease. Dua and Graff organized the UCI Machine Learning Repository, which supplies datasets about thyroid disease used across the literature. Tyagi and colleagues built a system that predicts thyroid disease using machine-learning techniques. Ahmad and colleagues proposed a decision-support system that merges ANFIS KNN and information gain to forecast thyroid disease. Chandel and colleagues conducted one of the comparisons using KNN and Naïve Bayes classifiers for thyroid disease. Ionita and Ionita applied data-mining methods to predict thyroid disease.

III. BASIC CONCEPTS AND BACKGROUND

The thyroid gland makes the hormones T4 and T3 while the pituitary gland sends out TSH to tell the thyroid what to do. When someone has hypothyroidism the thyroid does not work properly so T4 and T3 levels go down. Tsh goes up; when someone has hyperthyroidism the opposite happens with T4 and T3 going up and TSH going down. A clear study showed that hypothyroid patients had a TSH of 12.6 ± 5.4 mIU/L and

a T4 of 3.8 ± 1.2 µg/dL whereas hyperthyroid patients had a TSH of 0.1 ± 0.05 /L and a T4 of 18.5 ± 4.6 µg/dL showing why T4, T3 and TSH are helpful for diagnosis. Computer studies on thyroid disease often look at factors—age, gender, TSH, T3, T4, Free T4, T4U, FTI, medication use, symptoms, medical history and demographics—and feature selection can cut the list from 26 to 16 important factors while keeping the model strong. Artificial intelligence in health care uses machine learning algorithms such as Decision Trees, Random Forest, Logistic Regression, Naïve Bayes Support Vector Machines, K Nearest Neighbors, Gradient Boosting, XGBoost LightGBM and Neural Networks to predict disease classify conditions, assess risk monitor patients and read records. Even though these models are accurate they can act as "boxes" so explainability tools, like SHAP and LIME are used more and more to explain how each input— TSH, T3, TT4 and FTI—helps make a prediction.

IV. EXISTING TECHNIQUES AND METHODS

The UCI thyroid dataset has become a cornerstone in thyroid prediction research. Dataset is far from the only source. Researchers also use hospital records, thyroid datasets, Kaggle collections and other publicly available repositories. Some studies merge datasets. For example one study combined three sources. Created a set of 3,221 patient records with 30 attributes. A recent comparative analysis worked with 9,172 observations. Because these datasets vary widely in numbers, feature count and type disease categories missing data patterns and class distributions it is difficult to compare accuracy figures across studies directly. This makes preprocessing a step, in any thyroid AI pipeline. Typical preprocessing workflows include handling missing values dropping highly incomplete attributes encoding categorical variables managing outliers, scaling or normalizing features performing feature selection, balancing classes and splitting the dataset into training and testing sets. Many reviewed studies ultimately. Evaluate models and some even deploy them as web applications.

Feature selection is very important in this process because it helps find the useful variables and get rid of ones that don't add much value. This can make things faster. Sometimes help models work better when they see new data. In the studies that were looked at many different ways of selecting features were used. These include Featurewiz, SelectKBest elimination, backward elimination, bidirectional elimination, Recursive Feature Elimination, Principal Component Analysis, tree-based importance measures and statistical methods like Information Gain and Chi-square tests. One study did a comparison of forward, backward, bidirectional and Extra Trees methods for feature selection. A study from 2023, about AI showed that for most types of classifiers using fewer input features actually made things work better. This shows how important it is to choose features when trying to predict thyroid disease.

Among machine learning algorithms Decision Tree classifiers work by splitting the data using a series of if then rules, which makes them easy to understand but single trees can overfit the training data. Random Forest solves this problem by building decision trees and combining their predictions and it has shown especially good results for structured tabular healthcare data; one study found 99.45% accuracy after using Featurewiz for feature selection while another study achieved 91.42% accuracy with feature-importance based selection. Support Vector Machines (SVM) search for decision boundaries between classes and can work well on smaller datasets when their parameters are carefully adjusted whereas K-Nearest Neighbors (KNN) gives a class based on the labels of nearby points with results strongly influenced by the choice of distance metric feature scaling and the value of K. Logistic Regression offers a simple and clear baseline model while Naïve Bayes depends on the idea that features are independent—an idea that often is not true for thyroid-related variables and at least one study showed its performance was not good, for that reason.

Ensemble learning takes the concept of combining models further using techniques like bagging, boosting, Random Forest, XGBoost LightGBM, CatBoost, stacking and voting to make things more reliable and more accurate. For example a 2025 study looked at 17 machine learning models and used a hard-voting ensemble along with random oversampling to deal with class imbalance. A 2026 study showed that stacking and bagging ensembles did better than any model. A stacking setup that used XGBoost as the meta-learner got an F1-score of 0.9944. Deep learning methods—including Artificial Neural Networks (ANN) Deep Neural Networks (DNN) Convolutional Neural Networks (CNN) Long Short-Term Memory (LSTM) networks and hybrid CNN-LSTM models—have also been used in thyroid prediction. In one study comparing models on a Sick-euthyroid dataset ANN got an accuracy of 0.9587. This was just a little better than CatBoost with 0.9538 and XGBoost, with 0.9533. For structured and small medical datasets traditional tree-based ensemble models often work just as well or even better than deep learning methods.

Bio-inspired optimization algorithms have been used to help pick the features and set the right hyperparameters in thyroid disease prediction models. For example a hybrid method that joins Particle Snake Swarm Optimization with Random Forest called PSSO-RF reached 98.7% accuracy. It also had an F1-score of 98.47%, precision of 98.51% recall of 98.7% and specificity of 98%. Similar improvements were seen in Decision Tree and SVM models after using PSSO-based optimization.

At the time researchers have tried using GANs to create more data for small groups especially for rare thyroid conditions. One study used GAN-generated samples, with Kolmogorov–Arnold Networks (KAN). The KAN model hit 98.68% accuracy. On the expanded dataset Random Forest got an F1-score of 98.00%. The study also used SHAP and LIME to explain the model's predictions making them easier for doctors to understand.

V. COMPARATIVE ANALYSIS OF STUDIES

In below Table 1 which lists selected studies that investigate thyroid disorder detection using analysis, conventional machine learning, ensemble learning, optimization and explainable AI. These studies show promising classification results. For example one study used a stacking model with an XGBoost meta-learner. Another study used an XGBoost approach that included feature selection and class balancing. Some studies focused on feature-importance methods.

Used explainability tools such as SHAP and LIME. But the performance reported in these studies changes depending on the dataset characteristics the target definitions, the class distributions, the preprocessing methods and the validation protocols.

Table 1: Summary of Existing studies on Thyroid Disorder Detection

Study	Main Techniques	Reported Performance
Zhong & Wang [1]	machine learning. LR, DT, RF, SVM, XGBoost, LightGBM, ANN, DNN, stacking/bagging	Stacking with an XGBoost meta-learner: F1 = 0.9944
Ali et al. [2]	Pearson correlation, t-test	r = −0.82 (hypo), −0.78 (hyper)
Islam et al. [3]	SelectKBest, ensemble , ROS	99.72% accuracy; 100% sensitivity
Kesavulu & Kannadasan [4]	PSSO, RF/DT/SVM/KNN	PSSO-RF: 98.7%
Topşir et al. [5]	GAN,KAN, SHAP/LIME	KAN: 98.68%; RF F1: 98.00%
Hossain et al. [6]	DT, RF, GB, NB, KNN, LR, SVM + SHAP/LIME	RF: 91.42% (feature-importance set)
Lim et al. [7]	Featurewiz + ML	RF: 99.45%, F1: 0.99, AUC: 0.97

Also high classification accuracy does not always indicate that thyroid disorder detection can predict disease onset. Therefore I believe that standardized evaluation independent validation and longitudinal patient data are essential for building models, for early thyroid disorder risk prediction.

Table 2 compares the classification performance of seven machine learning algorithms. The comparison uses all features. The comparison also uses feature-importance selection. The comparison also uses feature selection. Based on the values that are reported feature-importance selection leads to performance. The better performance is seen for Decision Tree. The better performance is seen for Random Forest. The better performance is seen for Gradient Boosting. The better performance is seen for Naïve Bayes. The better performance is seen for K-Nearest Neighbors. Random Forest has the value in the feature-importance selection column. The highest value for Random Forest is 91.42 percent.
Logistic Regression has its value when all features are used. The highest value, for Logistic Regression is 84.48 percent.
These results show that feature selection can affect algorithms in ways. The comparison is only correct. The comparison is only correct when the same performance metric is used. The comparison is only correct when the same dataset is used. The comparison is only correct when the same target definition is used. The comparison is only correct when the same validation protocol is used.

Table 2: Comparison of Classification Performance Using Different Feature Selection Strategies

Algorithm	All Features	Feature Importance Selection	Univariate Selection
Decision Tree	82.90%	90.43%	89.55%
Random Forest	74.40%	91.42% (best)	90.40%
Gradient Boosting	83.97%	90.50%	89.35%
Naïve Bayes	16.44%	67.86%	56.30%
KNN	72.18%	86.22%	86.07%
Logistic Regression	84.48% (best)	73.15%	71.82%
SVM	84.38%	73.70%	74.15%

Feature selection needs to happen during the training process to avoid data leakage. More testing using cross-validation is needed. More testing using datasets is needed. More testing is needed to check how well the results work in general.

VI. APPLICATIONS

AI-based thyroid prediction has useful roles in healthcare. These models look at data like lab results and medical history to help find thyroid problems early. They can tell the difference between thyroid function, underactive thyroid, overactive thyroid and other conditions even breaking them down into more detailed groups. By putting together blood test numbers and personal info like age and gender the models can give a risk score. This helps doctors guess which thyroid condition a person might have. Tools like SHAP and LIME help explain how the model made its choice. They point out which lab values matter most. Things like TSH, T3, T4, TT4, T4U and FTI. Beyond diagnosis these systems can find patterns. They can show how lab results, age, sex and other factors link to thyroid diseases. Some studies have even built tools or fast-response systems that give predictions in real time. These tools could help clinics use tests smarter and cut down on time spent checking data by hand. Still before we use these tools widely we need testing, in real medical settings. We must be sure they work well and safely for patients.

VII. CHALLENGES AND LIMITATIONS

Although AI-based thyroid prediction has shown results there are still problems that stop it from being used in real hospitals. Many studies use single-source or very specific data sets like the UCI thyroid dataset, which might not show the variety of hospitals and patients. Class imbalance is also an issue; in some data sets the ratio is as uneven as 1135 to 75 which can make models favor the most common class even when using techniques like SMOTE, oversampling and GAN-based data

augmentation. Missing values, extra features and differences between hospitals and labs also make it harder to build and compare models. There is a missing piece when it comes to testing outside the original data and using many different centers. For example one study found that Random Forest accuracy dropped from 98.7% on UCI data to 95.80% on a data set, which shows a risk of overfitting. Also most data sets are not term so they don't give much information about how thyroid diseases change over time. Even though SHAP and LIME can help make models easier to understand their explanations don't always match what doctors consider important. So high accuracy alone is not enough to say the model is ready, for life; it also needs to be tested for safety, accuracy and being able to repeat the results.

VIII. RESEARCH GAPS

A synthesis of the reviewed literature highlights important research gaps. Although recent ensemble models often report high F1-scores most evaluations still focus primarily on accuracy and F1-score with less attention given to calibration, fairness and clinical usefulness [1]. Studies examining correlations have identified strong relationships, among thyroid hormones but computational Models rarely use TSH–T4 patterns to predict disease progression [2]. That's a gap. These patterns could give insight into how the disease changes over time but most models don't look at them. Similarly methods for dealing with class imbalance aren't strong or reliable enough yet. Some studies show promise with oversampling and ensemble methods [3]. Those results haven't been proven to work across different datasets or situations. It's hard to trust them Explainability tools like SHAP and LIME are used here and there. Not in a consistent way. Often they're added after the model is built, not from the start of the prediction pipeline [6]. That makes it harder to understand how the model is making decisions. Feature-selection tools such as Featurewiz have helped improve model performance in some cases. Those improvements are rarely tested on new independent data [7]. Without that we can't be sure the gains will hold up in real-world settings. Finally most research still depends on single-source datasets, like the UCI dataset [16]. These are helpful. Limited. We need more diverse multicenter clinical datasets that include multiple types of data. Then can the field move forward.

Table 3 shows research gaps and ideas for work in machine learning for predicting thyroid disease. Even though current studies have classification results more research is needed to make sure models are reliable, accurate, fair and work well in different situations. Looking at records over time could help study future disease risk. Comparing ways to balance classes and select features could make prediction models stronger. Explaining how models make decisions should be part of evaluating them to see if the predictions are based on features that're stable and important in medicine.

Also combining lab results with age, health history and lifestyle details could give an idea of patient risk. These gaps show the need, for a model that's easy to understand and tested on its own for predicting thyroid problems early.

Table 3: Research gap and Future Research Opportunities in Machine Learning Based Thyroid Disease Prediction

Sr.No	Description	Research Opportunity
1	Accuracy/-focused evaluation, with limited Attention to calibration and fairness[1]	Report calibration, fairness, and clinical-usefulness metrics
2	Static rather than longitudinal prediction[2]	Model repeated TSH/T3/T4 measurements over time
3	Incomplete class-imbalance solutions[3]	Develop robust GAN/SMOTE-based balancing strategies
4	Inconsistent explainability integration[6]	Embed SHAP/LIME throughout the full pipeline
5	Limited external/independent validation of feature-selection gains[7]	Validate feature-selection methods on independent, multi-hospital datasets
6	Limited multimodal data integration[9]	Combine laboratory, symptom, demographic, and lifestyle data
7	Dependence on public, single-source datasets[16]	Use large, diverse, real-world multi-center clinical datasets

IX. FUTURE DIRECTIONS

Future research should focus on creating integrated AI systems that combine data preprocessing, feature selection, class balancing, ensemble or deep-learning methods and explainability into a unified pipeline. Using repeated measurements of TSH, T3 and T4 over time could help researchers better understand how thyroid disease progresses and how patients respond to treatment. Including laboratory results alongside data, symptoms, medications and lifestyle factors may offer a more complete view of each patient's condition. Further studies should explore techniques to handle class imbalance, such as SMOTE, ADASYN and GAN-based data augmentation. Researchers should also develop feature-selection methods that're consistent, transparent and easy to reproduce across different studies. Explainability must be built into these systems from the start not added by combining tools like SHAP and LIME with counterfactual explanations that are clear, understandable and meaningful in a clinical setting. Finally these models need to be tested through validation, multi-center trials and prospective studies. Privacy-preserving techniques, such as learning will be

essential to allow collaboration, between institutions without requiring the sharing of sensitive patient data. Evaluation standards should go beyond accuracy. Include calibration, fairness and overall reliability. Together these efforts can lead to decision-support systems that're trustworthy, accurate and ready for real-world clinical use.

X. CONCLUSION

Artificial intelligence (AI) and machine learning (ML) are becoming more and more important in the study, prediction and detection of thyroid diseases and disorders. The research shows a change from old machine-learning methods to better ways, such as careful feature selection, group models and stacking models natural-inspired optimization deep learning, fake data creation and clear AI. Among the techniques studied Random Forest and XGBoost often show results. Group models can make predictions better with some research showing accuracy over 98–99% and F1-scores 0.99 especially after using feature-selection and balancing classes. These very high numbers should be looked at carefully. Differences in data how it is prepared and how it is tested make it hard to compare results between studies. Also many models have problems with data, classes that are not balanced no testing, on outside data depending on old patient records and not enough proof that they work in real hospitals. Future work should focus less on accuracy and more on making strong clear tested, long-term and mixed models. In the end these tools should help doctors make decisions not take over the job of diagnosing patients.

CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest.

REFERENCES

[1] J. Zhong and Y. Wang, "A comparative empirical analysis for robust thyroid disorder detection using machine learning techniques," Journal of Electrical Systems and Information Technology, vol. 13, Art. no. 12, 2026. Available from: https://doi.org/10.1186/s43067-026-00315-3

[2] H. Ali, S. Ullah, S. Ullah, R. Ahmad, Q. Ali, S. T. Shah, and S. Johar, "Evaluating the relationship between TSH and T4 in thyroid dysfunction," Journal of Medical & Health Sciences Review, vol. 2, no. 3, 2025. Available from: https://doi.org/10.62019/wtj65f49

[3] M. R. Islam, A. Islam Chowdhury, S. Shama, and M. M. H. Lamyea, "Enhanced thyroid disease prediction using ensemble machine learning: A high-accuracy approach with feature selection and class balancing," Discover Artificial Intelligence, vol. 5, Art. no. 9, 2025. Available from: https://doi.org/10.1007/s44163-025-00225-9

[4] D. Kesavulu and R. Kannadasan, "Improved bio-inspired with machine learning computing approach for thyroid prediction," Scientific Reports, vol. 15, Art. no. 22524, 2025. Available from: https://doi.org/10.1038/s41598-025-03299-8

[5] A. Topşir, F. Güler, E. Çetin, M. F. Burak, and M. Ağraz, "Thyroid disease classification using generative adversarial networks and Kolmogorov-Arnold network for three-class classification," BMC Medical Informatics and Decision Making, vol. 25, Art. no. 284, 2025. Available from: https://doi.org/10.1186/s12911-025-03014-7

[6] M. B. Hossain, A. Shama, A. Adhikary, A. D. Raha, K. M. A. Uddin, M. A. Hossain, I. Islam, S. A. Murad, M. S. Munir, and A. K. Bairagi, "An explainable artificial intelligence framework for the predictive analysis of hypo and hyper thyroidism using machine learning algorithms," Human-Centric Intelligent Systems, vol. 3, pp. 211–231, 2023. Available from: https://doi.org/10.1007/s44230-023-00027-1

[7] S. T. Lim, K. W. Khaw, X. Y. Chew, and W. C. Yeong, "Prediction of thyroid disease using machine learning approaches and Featurewiz selection," Journal of Telecommunication, Electronic and Computer Engineering, vol. 15, no. 3, 2023. Available from: https://doi.org/10.54554/jtec.2023.15.03.002

[8] T. Alyas, M. Hamid, K. Alissa, T. Faiz, N. Tabassum, and A. Ahmad, "Empirical method for thyroid disease classification using a machine learning approach," BioMed Research International, vol. 2022, Art. no. 9809932, 2022. Available from: https://doi.org/10.1155/2022/9809932

[9] R. Chaganti, F. Rustam, I. De La Torre Díez, J. L. V. Mazón, C. L. Rodríguez, and I. Ashraf, "Thyroid disease prediction using selective features and machine learning techniques," Cancers, vol. 14, no. 16, Art. no. 3914, 2022. Available from: https://doi.org/10.3390/cancers14163914

[10] S. S. Islam, M. S. Haque, M. S. U. Miah, T. B. Sarwar, and R. Nugraha, "Application of machine learning algorithms to predict the thyroid disease risk: An experimental comparative study," PeerJ Computer Science, vol. 8, Art. no. e898, 2022. Available from: https://doi.org/10.7717/peerj-cs.898

[11] L. Aversano, M. L. Bernardi, M. Cimitile, M. Iammarino, P. E. Macchia, I. C. Nettore, and C. Verdone, "Thyroid disease treatment prediction with machine learning approaches," Procedia Computer Science, vol. 192, pp. 1031–1040, 2021. Available from: https://doi.org/10.1016/j.procs.2021.08.106

[12] H. A. U. Rehman, C. Y. Lin, Z. Mushtaq, and S. F. Su, "Performance analysis of machine learning algorithms for thyroid disease," Arabian Journal for Science and Engineering, vol. 46, no. 10, pp. 9437–9449, 2021. Available from: https://doi.org/10.1007/s13369-020-05206-x

[13] K. Salman and E. Sonuç, "Thyroid disease classification using machine learning algorithms," Journal of Physics: Conference Series, vol. 1963, no. 1, Art. no. 012140, 2021. Available from: https://doi.org/10.1088/1742-6596/1963/1/012140

[14] U. Pawar, D. O'Shea, S. Rea, and R. O'Reilly, "Incorporating explainable artificial intelligence (XAI) to aid the understanding of machine learning in the healthcare domain," CEUR Workshop Proceedings, vol. 2771, pp. 169–180, 2020. Available from: https://ceur-ws.org/Vol-2771/

[15] A. Begum and A. Parkavi, "Prediction of thyroid disease using data mining techniques," in Proceedings of the 5th International Conference on Advanced Computing & Communication Systems (ICACCS), 2019, pp. 342–345. Available from: https://doi.org/10.1109/ICACCS.2019.8728320

[16] D. Dua and C. Graff, UCI Machine Learning Repository. Irvine, CA, USA: University of California, Irvine, School of Information and Computer Sciences, 2019. Available from: https://doi.org/10.24432/C5D010

[17] A. Tyagi, R. Mehra, and A. Saxena, "Interactive thyroid disease prediction system using machine learning technique," in Proceedings of the 5th IEEE International Conference on Parallel, Distributed and Grid Computing (PDGC), 2018, pp. 689–693. Available from: https://doi.org/10.1109/PDGC.2018.8745910

[18] W. Ahmad, L. Huang, A. Ahmad, F. Shah, A. Iqbal, and A. Saeed, "Thyroid diseases forecasting using a hybrid decision support system based on ANFIS, k-NN and information gain method," Journal of Applied Environmental and Biological Sciences, vol. 7, no. 10, pp. 78–85, 2017. Available from: https://www.researchgate.net/publication/320350249_Thyroid_Diseases_Forecasting_Using_a_Hybrid_Decision_Support_System_Based_on_ANFIS_k-NN_and_Information_Gain_Method

[19] K. Chandel, V. Kunwar, S. Sabitha, T. Choudhury, and S. Mukherjee, "A comparative study on thyroid disease detection

using K-nearest neighbor and Naive Bayes classification techniques," CSI Transactions on ICT, vol. 4, no. 2, pp. 313–319, 2016. Available from: https://doi.org/10.1007/s40012-016-0100-5

[20] I. Ionită and L. Ionită, "Prediction of thyroid disease using data mining techniques," BRAIN. Broad Research in Artificial Intelligence and Neuroscience, vol. 7, no. 3, pp. 115–124, 2016. Available from: https://editora.ro/index.php/brain/article/view/211

ABOUT THE AUTHORS



Roshani Koli is a researcher in the field of Computer Science and Information Technology. She is currently pursuing her Ph.D. She completed her Master of Computer Applications (MCA) from Sarvajanik University in 2024 and her Bachelor of Commerce (B.Com.) from Veer Narmad South Gujarat University in 2022. Her academic background reflects an interdisciplinary foundation combining computing and commerce.



Dr. Hemangini Patel specializes in Computer Science and Information Technology. She received her Ph.D. in Computer Science from Veer Narmad South Gujarat University in 2018. She completed her Master of Science in Information Technology (M.Sc. IT) in 2005 and Bachelor of Science in Information Technology (B.Sc. IT) in 2003 from Veer Narmad South Gujarat University. Her research interests include Information Retrieval, Web Mining, Machine Learning, and Deep Learning. She is also a recognized Ph.D./M.Phil. guide and postgraduate teacher.